

Oorzaak en Gevolg

Vakgroep Econometrie, Faculteit der Economische Wetenschappen en Econometrie
Vrije Universiteit Amsterdam

Rede uitgesproken op 4 april 1995 bij het aanvaarden van het ambt van hoogleraar
in de econometrie.

Geert Ridder

1 Oorzaak en Gevolg

Omdat een oratie, zeker aan de Vrije Universiteit, een preek is, wil ik met een schriftlezing beginnen. De schriftlezing is, zoals bekend, geen onderdeel van de preek en ik reken er dan ook op dat de rector mij wat extra tijd geeft. De lezing is Richteren 16:4-22, het verhaal van Simson en Delila.

Daarna vatte hij liefde op voor een vrouw in het dal Sorek, Delila genaamd. En de stadsvorsten der Filistijnen kwamen bij haar en zeiden tot haar: Tracht door overreding van hem te weten te komen, waardoor zijn kracht zo groot is, en hoe wij hem kunnen overmeesteren en binden om hem te bedwingen. Wij zullen u dan ieder elfhonderd zilverstukken geven. Toen zeide Delila tot Simson: Vertel mij toch, waardoor uw kracht zo groot is, en waarmee gij gebonden zoudt moeten worden om u te bedwingen. En Simson zeide tot haar: Indien men mij bindt met zeven verse pezen, die nog niet verdroogd zijn, dan zal ik machteloos wezen en gelijk aan ieder ander mens. Toen brachten de stadsvorsten der Filistijnen haar zeven verse pezen, die nog niet verdroogd waren, en zij bond hem daarmee, terwijl zij in haar binnenvertrek in hinderlaag lagen. Toen riep zij: De Filistijnen over u, Simson! Maar hij verscheurde de pezen, zoals een vlassnoer verscheurd wordt, wanneer het vuur geroken heeft. En het werd niet bekend waarin zijn kracht lag.

Delila zeide tot Simson: Zie, gij hebt mij bedrogen en mij leugens verteld. Zeg mij nu toch, waarmee gij gebonden kunt worden. En hij zeide tot haar: Indien men mij stevig bindt met nieuwe touwen, die nog niet gebruikt zijn, dan zal ik machteloos wezen en gelijk aan ieder ander mens. Daarop nam Delila nieuwe touwen, bond hem daarmee en riep hem toe: De Filistijnen over u, Simson!-er lagen in het binnenvertrek mannen in hinderlaag-maar hij scheurde ze van zijn armen af als een draad.

Delila zeide tot Simson: Tot nu toe hebt gij mij bedrogen en mij leugens verteld. Zeg mij, waarmee gij gebonden kunt worden. Toen zeide hij tot haar: Als gij de zeven haarvlechten van mijn hoofd met de schering van een weefgetouw samenweeft. En zij hechtte ze vast met een pin, waarna zij hem toeriep: De Filistijnen over u, Simson! Maar, toen hij uit zijn slaap ontwaakte, rukte hij de weverspin en de schering los.

Zij zeide tot hem: Hoe kunt gij nog zeggen: Ik heb u lief, terwijl uw hart mij niet toebehoort? Nu hebt gij mij reeds driemaal bedrogen en mij niet verteld, waardoor uw kracht zo groot is. En toen gebeurde het, nadat zij dag aan dag bij hem met haar vragen was blijven aandringen en aanhouden, dat hij ongeduldig werd tot stervens toe, zijn gehele hart voor haar blootlegde en haar zeide: Geen scheermes is ooit op mijn hoofd gekomen, want van mijn moederschoot af ben ik een nazireeër Gods. Indien ik geschoren werd, zou mijn kracht van mij wijken, en ik zou machteloos wezen en gelijk aan ieder ander mens. Toen Delila zag, dat hij zijn gehele hart voor haar blootgelegd had, liet zij de stadsvorsten der Filistijnen roepen, en zeide: Ditmaal moet gij komen, want hij heeft zijn gehele hart blootgelegd. En de stadsvorsten der Filistijnen kwamen bij haar en brachten het geld mee. Daarop liet zij hem op haar knieën inslapen, riep iemand en liet de zeven vlechten van zijn hoofd afscheren. Zo begon zij hem in bedwang te krijgen, want zijn kracht week van hem. En zij riep: De Filistijnen over u, Simson! Toen ontwaakte hij uit zijn slaap en dacht: evenals de vorige keren zal ik vrijkomen en mij losrukken; maar hij wist niet, dat de Here van hem geweken was. De Filistijnen

grepen hem, staken hem de ogen uit, voerden hem naar Gaza en boeiden hem met twee koperen ketenen. En hij moest in de gevangenis de molen draaien. Maar van het ogenblik af, dat zijn hoofdhaar afgeschoren was, begon het weer aan te groeien.

Dit voorbeeld van een bijbels experiment is een goed begin van een discussie van causaliteit in de economische wetenschap. De economische wetenschap pretendeert iets te zeggen over de economische werkelijkheid en een confrontatie van uit de theorie afgeleide wetten met die werkelijkheid is daarom gewenst. De theorie kan voorspellen, dat een stijging van de prijs van koffie, bij een gelijkblijvend inkomen en andere omstandigheden, zal leiden tot een daling van de vraag, maar alleen analyse van gegevens kan ons leren of dit inderdaad het geval is en in welke mate. Om die reden maken economen al een eeuw gebruik van statistische methoden om verbanden tussen economische grootheden te onderzoeken. Eerst met grafieken, later met technieken als correlatie- en regressierekening. Deze activiteiten waren in de jaren 30 zo omvangrijk geworden, dat in 1931 besloten werd de Econometric Society op te richten die tot doel had het gebruik van kwantitatieve methoden in de economische wetenschap te bevorderen. De econometrie werd een zelfstandig vak en de door econometristen ontwikkelde methoden werden de werktuigen waarmee empirische economen de economische werkelijkheid analyseerden.¹

De econometrische methoden werden in de loop der jaren verfijnder. Een belangrijke stap was de introductie van de waarschijnlijkheidsrekening bij de beschrijving van economische relaties.² De volgtijdelijke waarden van economische grootheden werden opgevat als de uitkomsten van een kansexperiment. De analyse van kansexperimenten met behulp van uitkomsten is het gebied van de wiskundige statistiek en deze stap maakte het mogelijk het gehele arsenaal van wiskundig-statistische methoden te gebruiken bij de analyse van economische wetten. Dat betekende wel, dat de kansexperimenten met een wiskundig model beschreven moesten worden. Econometristen gebruikten de deterministische relaties tussen economische grootheden, zoals die voorspeld werden door economische theorieën, en vatten de afwijkingen tussen de waarnemingen en de relaties op als kansvariabelen, de storingen, die de invloed van alle niet opgenomen factoren weerspiegelden.

De moderne econometrie is opgetrokken op dit kanstheoretische fundament. De grote aantrekkingskracht van dit fundament is, dat het moeilijke problemen, zoals de keuze van een model, hanteerbaar maakt. Daarbij kan de econometrist kiezen tussen verschillende benaderingen. Hij³ kan gebruik maken van de toetstheorie, zoals ontwikkeld door Neyman en Pearson en verfijnd tot een hulpmiddel bij de modelselectie door Hendry⁴ of hij kan de Bayesiaanse aanpak volgen waarin kansen toegekend worden aan mogelijke modellen. Als de gegevens verzameld zijn, begint in beide benaderingen een zoektocht in de wereld der modellen en afhankelijk van de gebruikte criteria wordt uiteindelijk een model gepresenteerd als de voorlopig beste beschrijving van de werkelijkheid. De econometrist schat en toetst of berekent posteriors op basis van de verzamelde gegevens.

Wat heeft dat nu met oorzaak en gevolg te maken, zult u zich afvragen. Het antwoord is, dat het daar weinig mee te maken heeft. De vraag of de relaties die op de boven beschreven manier gemeten worden, als oorzakelijke verbanden geïnterpreteerd kunnen worden, is een van de oudste en hardnekkigste vragen uit de econometrie. Morgan noemt het op verschillende plaatsen in haar boek over de geschiedenis van de econometrie (Morgan, 1990) het fundamentele probleem van de econometrie. Keynes noemt het in zijn bekende kritiek op het werk van Tinbergen.⁵ Ook mod-

¹(Morgan, 1990) geeft een overzicht van de ontwikkeling van de econometrie.

²Dit gebeurde door Haavelmo (Haavelmo, 1944).

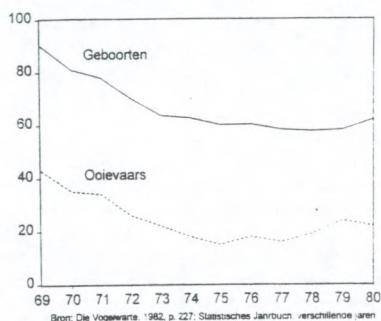
³In het vervolg zal ik hij in generieke zin gebruiken.

⁴Zie zijn leerboek (Hendry, 1995)

⁵Zie (Keynes, 1939). Morgan plaatst de kritiek van Keynes in perspectief.

erne critici van econometrische methoden verwijzen naar dit probleem.⁶ Ondanks deze kritiek komt het probleem in leerboeken nauwelijks aan de orde. Ik vermoed dat de reden hiervoor is, dat de leerboeken aansluiten bij de traditie dat voor de econometrist maar twee dingen relevant zijn: de gegevens en de klasse van modellen die de gegevens kunnen beschrijven. Ik zal in het vervolg betogen, dat voor causale uitspraken meer informatie nodig is. Overigens is het probleem van de causale inferentie niet beperkt tot de econometrie. In de oratie van Bouter (Bouter,1993) die als hoogleraar epidemiologie aan de Faculteit der Geneeskunde van de VU verbonden is, werd het probleem ook genoemd en de statisticus Freedman voert al jaren een kruistocht tegen causale claims in de sociale wetenschappen.⁷

Het probleem van de causale inferentie wordt geïllustreerd door de volgende figuur waarin het verloop van geboorten in West-Duitsland en broedparen ooievaars in Baden-Württemberg aangegeven wordt.⁸



Figuur 1: Broedparen Ooievaars in Baden-Württemberg en Geboorten ($\times 10000$) in West-Duitsland: 1969-80

Het gesuggereerd verband zal door weinigen als causaal beschouwd worden. Toch is het niet eenvoudig om de relatie als een schijnverband te ontmaskeren. In andere woorden: het is gevaarlijk om uit een correlatie te concluderen dat er een oorzakelijk verband bestaat.⁹

2 Een Model voor Causale Inferentie

De vraag wat een oorzakelijk verband is en hoe je het kunt vaststellen, is een vraag die al in de Griekse oudheid aan de orde gesteld werd. Ik laat de discussie over verschillende soorten oorzaken en hun mogelijke werking graag aan filosofen over. Ik kies in het volgende een praktisch uitgangspunt, dat een groot aantal, maar niet

⁶Bijvoorbeeld Black (Black,1982) die zijn artikel besluit met de opmerking: "I doubt, however, that traditional econometric methods will survive."

⁷Zie bijvoorbeeld (Freedman,1994). Het boek van Wang (Wang,1993) bevat een aardige, zij het af en toe wat hysterisch geformuleerde bespreking van het probleem.

⁸Een soortgelijke figuur, zij het met de broedparen ooievaars in West-Duitsland staat in (Sies,1988).

⁹De econometrist is niet geheel machteloos. Schijnregressies tussen niet-stationaire tijdreeksen kunnen opgespoord worden door de residuen van de regressie te bestuderen. Als er het vermoeden bestaat dat een derde variabele, die met de beschouwde variabelen samenhangt, verantwoordelijk is voor het verband en die variabele gemeten is, dan zal het opnemen van die derde variabele in de relatie de schijnrelatie ongedaan maken.

alle causale uitspraken verheldert. Dit 'model' voor causale inferentie is bedacht door Rubin (Rubin, 1974) en uitgewerkt door Holland (Holland, 1986). Ik zal het dan ook het Rubin-Holland (RH) model noemen.

Het RH model bespreekt causale inferentie in de context van experimenten. Een experiment kan een echt experiment zijn of een mogelijk experiment, dat echter niet uitgevoerd is. In de economie hebben we meestal met deze tweede vorm van experimenten te maken. Een belangrijke implicatie van de beperking tot experimenten is, dat we de gevolgen van een oorzaak kunnen onderzoeken, maar niet de oorzaak van een gevolg. Immers in een experiment staat vast wat de oorzaak is en wat het gevolg. In werkelijkheid is dat niet altijd duidelijk. Experimenten zijn niet geschikt om de oorzaak van een gevolg te achterhalen en zoals we aan het einde van mijn betoog zullen zien heeft deze fout een groot aantal Filistijnen de kop gekost. Deze beperking verlost ons van vele moeilijke filosofische problemen.

Meestal maken we de indeling in oorzaak en gevolg op basis van de volgorde waarin de gebeurtenissen optreden: de oorzaak gaat vooraf aan het gevolg. In de economie hoeft dat niet zo te zijn. Een voorbeeld is, dat werklozen intensiever gaan zoeken als het moment waarop de werkloosheidsuitkering omlaag gaat nadert.¹⁰ Het intensievere zoekgedrag kan echter moeilijk als oorzaak van de daling van de uitkering beschouwd worden. Economische agenten kijken vooruit en dat kan leiden tot een omkering van de volgorde waarin oorzaak en gevolg optreden. De indeling in oorzaak en gevolg is dan ook een keuze die gerechtvaardigd kan worden door theorie, maar niet door waarneming.

In een experiment is een oorzaak een relatief begrip. In het eenvoudigste geval, waartoe ik me in het vervolg zal beperken, vergelijken we voor iedere eenheid de respons met en de respons zonder de behandeling. Om een en ander te verduidelijken zal ik steeds het voorbeeld van de meting van het effect van een scholingsprogramma, dat de positie op de arbeidsmarkt van een bepaalde groep probeert te verbeteren, op het inkomen dat één jaar na afloop van de scholing verdiend wordt, bekijken. In dat geval is het causale effect van de scholing op individu i het verschil in inkomen van individu i met de scholing, Y_{si} en zonder de scholing, Y_{ci} (de c geeft de controlebehandeling aan)

$$Y_{si} - Y_{ci} \quad (1)$$

Het is duidelijk dat dit verschil, en dus het causale effect, alleen uitgerekend kan worden, als het denkbaar is dat een individu zowel s als c ondergaat. Als dat niet mogelijk is zonder iets essentieels aan het individu te veranderen, dan kan het causale effect niet vastgesteld worden. In het RH model kunnen, als de experimentele eenheid het individu is, geslacht en ras dan ook geen oorzaken zijn, een politiek-correcte conclusie.¹¹ Merk verder op, dat als we I individuen beschouwen, ieder individu mogelijk anders op de scholing reageert. Het causale effect verschilt van individu tot individu.

Als we het causale effect van scholing op individu i proberen te meten, dan stuiten we op een probleem. Het is immers onmogelijk om voor het individu zowel het inkomen met en zonder de scholing te meten. Dit is het *fundamentele probleem van de causale inferentie*: het is onmogelijk om voor i zowel Y_{si} als Y_{ci} te meten; we kunnen slechts één van beide meten.

Toch doen we in de praktijk, en in de wetenschap, causale uitspraken. Het is niet moeilijk in te zien, dat dit mogelijk is, als we tenminste bereid zijn veronderstellingen te maken. Als we in een auto zitten, dan lijkt het redelijk te veronderstellen, dat zolang we het sleuteltje niet omdraaien de auto niet start. Het feit dat we de

¹⁰Zie (Van den Berg, 1990).

¹¹Als de experimentele eenheid de baan is, d.w.z. als we een populatie van banen bekijken, dan kunnen ras en geslacht natuurlijk wel oorzaken zijn, als het tenminste denkbaar is, dat personen van beide geslachten de baan kunnen vervullen.

auto niet starten heeft ook geen gevolgen voor de uitkomst die we waarnemen als we het sleuteltje wel omdraaien. Natuurlijk moeten we niet te lang wachten, want dan is de accu leeg. Onder deze veronderstellingen kunnen we de uitspraak doen, dat het omdraaien van het sleuteltje de auto doet starten. Een tweede veronderstelling die vaak gemaakt wordt, is dat alle experimentele eenheden identiek zijn. Bij deze veronderstelling vergelijken we de gevolgen van het omdraaien van het sleuteltje in het contact van de ene auto met de gevolgen van het achterwege laten van deze handeling bij de andere. De veronderstellingen lijken zeer redelijk in het gekozen voorbeeld en dat is mooi, want ons model voor causale inferentie moet ook in huiselijke omstandigheden werken. De veronderstellingen gaan zeker niet op in het voorbeeld van het effect van scholing op inkomen. Immers het vergelijken van het inkomen vóór de scholing met dat na de scholing meet niet het causale effect van de scholing; in de periode zijn er behalve de scholing vele andere observeerbare en niet-observeerbare dingen veranderd. De tweede aanpak zou betekenen, dat we het individu vergelijken met een identiek individu. Dat is ook niet mogelijk. Overigens verklaart deze tweede veronderstelling de bijzondere aandacht van epidemiologen en sociale wetenschappers voor tweelingen. De Filistijnen gebruikten in hun experimenten op Simson de eenvoudige voor-na vergelijking.

Als de veronderstelling dat de beschouwde individuen identiek zijn of blijven niet opgaat, dan is het veel moeilijker om het causale effect te meten. We kunnen toch iets doen, indien we bereid zijn met minder dan het causale effect voor alle individuen i genoeg te nemen. We zullen slechts het *gemiddelde causale effect* voor een groep individuen kunnen meten. Een groep individuen noemen we in de statistiek een populatie. Het gemiddelde causale effect T voor een populatie van I individuen is

$$T = E(Y_s - Y_c) \quad (2)$$

In deze vergelijking geeft de operator E aan, dat het rekenkundig gemiddelde van I waarnemingen genomen is. In het vervolg zij alle verwachtingen rekenkundige gemiddelden en kansen de fractie van de populatie met een bepaalde eigenschap. De beperking tot het gemiddelde causale effect maakt het probleem misschien oplosbaar, maar zegt de oplossing ons nog iets? Het gemiddelde effect heeft als alle gemiddelden een ernstig defect: het is eigenlijk voor geen van de individuen in de populatie relevant. Andere grootheden lijken interessanter, bijvoorbeeld de fractie individuen waarvoor de scholing het inkomen verhoogt. Om die fractie te berekenen moeten we echter de *gezamenlijke* verdeling van Y_{si} en Y_{ci} bepalen. Als we even nadenken over het fundamentele probleem van de causale inferentie, dan zien we in dat dit onmogelijk is. We kunnen slechts hopen de marginale verdelingen van Y_{si} en Y_{ci} te achterhalen en het is bekend dat kennis van de marginale verdelingen niet veel leert over de gezamenlijke verdeling.¹² We kunnen natuurlijk in plaats van het gemiddelden de medianen van de verdelingen van Y_{si} en Y_{ci} vergelijken.

We kunnen de vergelijking voor het gemiddelde causale effect als volgt herschrijven

$$T = E(Y_s) - E(Y_c) \quad (3)$$

Hierbij maken we gebruik van de gelijkheid van het gemiddelde van een verschil aan het verschil van de gemiddelden. Deze vergelijking suggereert een oplossing voor het fundamentele probleem van de causale inferentie. We kunnen immers het gemiddelde inkomen met scholing en het gemiddelde inkomen zonder scholing schatten met het gemiddelde inkomen van de individuen die respectievelijk wel en geen scholing ontvangen hebben. Ik moet daarbij benadrukken, dat het gemiddelde inkomen met

¹²De Fréchet-grenzen (Fréchet, 1951) geven een onder- en bovengrens voor de verdelingsfunctie van de gezamenlijke verdeling.

scholing in de vergelijking berekend wordt over alle I individuen in de populatie, terwijl het gemiddelde inkomen van personen die scholing ontvangen hebben berekend wordt over de *deelpopulatie* die inderdaad aan de scholing deelgenomen heeft. De schatting van de niet-observeerbare gemiddelden uit de vergelijking door de observeerbare gemiddelden lijkt op het eerste gezicht een goed idee. Echter bij nadere beschouwing kleven er nogal wat bezwaren aan deze procedure. Als de deelname aan de scholing vrijwillig is en zich minder personen aanbieden dan geschoold kunnen worden, dan ligt de beslissing tot deelname geheel bij die personen. Zij kiezen voor scholing en doen dat, omdat zij er beter van denken te worden. Als zij bij hun beslissing het verwachte inkomen dat zij met en zonder scholing kunnen verwerven vergelijken en die verwachtingen redelijk nauwkeurig zijn, dan zullen personen met een relatief groot causaal effect van de scholing tot deelname besluiten. Als het inkomen zonder scholing ongeveer gelijk is voor deelnemers en niet-deelnemers, dan is het verschil tussen het gemiddelde inkomen van deelnemers en niet-deelnemers een overschatting van het gemiddelde causale effect in de populatie van I individuen.

Laten we ook eens het geval bekijken dat de capaciteit van de scholingscentra onvoldoende is. In dat geval zal er geselecteerd moeten worden. De uitkomst van die selectie hangt af van de vrijheid en de doelstelling van de uitvoerders van het scholingsprogramma. Als deze uitvoerders gebaat zijn bij succes, dat afgemeten wordt aan de gerealiseerde inkomensgroei, en vrij zijn bij de selectie, dan zullen de personen met een naar verwachting groot causaal effect van scholing en personen die het ook zonder scholing relatief goed zouden doen geselecteerd worden. Het gemeten effect zal onder deze veronderstellingen nog meer verschillen van het gemiddelde causale effect dan bij keuze door de potentiële deelnemers.

We kunnen de voorgaande discussie als volgt formaliseren. Zij S_i een indicatorvariabele die gelijk is aan s als i aan de scholing deelneemt en gelijk is aan c als dat niet zo is. Het waargenomen inkomen van i hangt af van deelname aan de scholing of niet en wordt gegeven door $Y_{S,i}$. Het *waargenomen* gemiddelde effect van de scholing is dan gelijk aan

$$\begin{aligned} T_W &= E(Y_S | S = s) - E(Y_S | S = c) \\ &= E(Y_s | S = s) - E(Y_c | S = c) \end{aligned} \quad (4)$$

Het waargenomen gemiddelde effect is dus gelijk aan het verschil van het gemiddelde inkomen van de geschoolden en de niet-geschoolden. We kunnen dit vertalen naar het gemiddelde causale effect. Er geldt

$$T_W = E(Y_s - Y_c | S = s) + (E(Y_c | S = s) - E(Y_c | S = c)) \quad (5)$$

Het waargenomen gemiddelde effect is dus de som van het gemiddelde causale effect voor de deelnemers en een term die opgevat kan worden als het verschil in uitgangspositie van deelnemers en niet-deelnemers, zij het dat die uitgangspositie betrekking heeft op de situatie na de scholing. Als de uitgangsposities van deelnemers en niet-deelnemers ongeveer gelijk zijn, dan meten we het gemiddelde causale effect voor de deelnemers. In het geval dat de deelnemers beslissen over deelname, zouden we hier tevreden mee kunnen zijn. Zekerheid over gelijke uitgangsposities hebben we echter zelden.¹³

Het waargenomen gemiddelde effect is dus gelijk aan het gemiddelde causale effect, zelfs niet als we ons beperken tot de deelnemers. Het probleem is, dat bij de keuze of de selectie voor deelname het (verwachte) resultaat van de scholing een rol speelt. In de jaren twintig van deze eeuw suggereerden statistici, dat een zuivere meting van het causale effect alleen mogelijk, als we ingrijpen in het proces dat de

¹³ Deze vergelijking is de basis van de verschil-in-verschil schatter die vaak bij de evaluatie van scholingsprogramma's gebruikt is. Zie (Ashenfelter, 1978).

deelname bepaalt.¹⁴ Een in sommige ogen wat Filistijnse oplossing is om de deelname door het lot te laten bepalen. We spreken dan van gerandomiseerde toewijzing. Bij gerandomiseerde toewijzing is er behalve de deelname gemiddeld geen verschil tussen deelnemers en niet-deelnemers. Helemaal zeker zijn we hiervan nooit, maar bij een voldoende groot aantal waarnemingen is dit een goede benadering. Een gevolg is, dat het gemiddelde inkomen van de niet-deelnemers een zuivere schatting is van het gemiddelde inkomen van de deelnemers, indien zij niet aan de scholing deelgenomen hadden, en dat het gemiddelde inkomen van de deelnemers een zuivere schatting is van dat van de niet-deelnemers, als zij geschoold zouden zijn. Dit betekent, dat het waargenomen gemiddelde causale effect T_W gelijk is aan het gemiddelde causale effect T . Met de eerder geïntroduceerde symbolen leidt gerandomiseerde toewijzing tot onafhankelijkheid van Y_s, Y_c van S . Gerandomiseerde toewijzing is standaard in het landbouwkundige en het medische onderzoek. Voordat ik op de toepassing in het economische onderzoek inga, wil ik nog twee opmerkingen maken over het RH model voor causale inferentie.

Het RH model lijkt sterk op modellen die econometristen al lange tijd gebruiken. Roy (Roy, 1951) formuleerde de verdeling van tijd tussen jagen en vissen op deze manier. Econometristen formaliseerden later dit model als het regime-regressiemodel met endogene regimekeuze. Een bijzonder geval van dit model is het steekproefselectiemodel van Heckman (Heckman, 1979). Het steekproefselectiemodel wordt veel gebruikt in het toegepaste onderzoek, omdat het een gemakkelijk implementeerbare oplossing van het fundamentele probleem van de causale inferentie belooft. Voor deze oplossing is het noodzakelijk om een model te maken van de deelname en de samenhang tussen deelname en uitkomst. Omdat we in de meeste gevallen lang niet alle determinanten van de uitkomst en de beslissing tot deelname kennen en het mogelijk is, dat de niet-observeerbare determinanten samenhangen, zijn we genoodzaakt veronderstellingen te maken over de gezamenlijke verdeling van niet-observeerbare grootheden. Dat is een griezelige bezigheid, want het kan betekenen dat de gevonden resultaten slechts een getrouwe weergave van de gemaakte veronderstellingen zijn.

Zoals ik eerder benadrukt heb, beperk ik me tot het meten van het gevolg van een oorzaak. Dat lijkt een grote beperking, want veel statistisch onderzoek in wetenschappen waar weinig theorie voorhanden is of waar de theorie gewantrouwd wordt, is juist gericht op het ontdekken van oorzaken van een gemeten gevolg. In zijn meest extreme vorm gebeurt dit in volautomatische procedures.¹⁵ Er is een lange traditie op dit gebied. Er is tegenwoordig zelfs een filosofische rechtvaardiging van een dergelijke aanpak, de probabilistische theorie van causaliteit.¹⁶ In deze theorie wordt gesteld dat X een oorzaak is van Y , als er geen Z gevonden kan worden die indien constant gehouden, X en Y onafhankelijk maakt. Ik denk, dat dit geen vruchtbare aanpak is, aangezien de praktische uitwerking van dit idee tot problemen leidt.¹⁷ Stel bijvoorbeeld, dat scholing alleen een hoger inkomen geeft door een kleinere kans op werkloosheid. In dat geval zou scholing geen oorzaak zijn van het hogere inkomen!

Ook het in de tijdreeksconometrie gebruikte begrip Granger causaliteit (Granger, 1969) gaat uit van deze aanpak. In het geval van twee tijdreeksen $\{X_t, t = 1, \dots, T\}$ en $\{Y_t, t = 1, \dots, T\}$ van de variabelen X en Y , veroorzaakt X Y , als $X_t, X_{t-1}, \dots$ een statistisch significante bijdrage hebben in de regressie van Y_t op $X_t, X_{t-1}, \dots, Y_{t-1}, Y_{t-2}, \dots$ ¹⁸ De variabele Z is hier dus de reeks van vertraagde waarden van Y . Het is mogelijk, dat

¹⁴ Dit inzicht werd eerst toegepast bij experimenten in de landbouw (Fisher, 1926).

¹⁵ Zoals het TETRAD algoritme (Spirtes et al., 1993).

¹⁶ Zie (Suppes, 1970; Glymour et al., 1993).

¹⁷ Zie ook (Freedman, 1994).

¹⁸ Dit correspondeert met Grangers definitie van onmiddellijke causaliteit. In de gewone definitie van Granger causaliteit wordt X_t weggelaten.

Verklar. var.	Afhankelijke variabele			
	<i>GEBOORT</i>	<i>OOIVRS</i>	<i>GEBOORT</i>	Δ <i>GEBOORT</i>
<i>OOIVRS</i>	.40 (.15)	.	1.17 (.12)	.
<i>GEBOORT</i>	.	.50 (.41)	.	.
<i>OOIVRS</i> ₋₁	.	.24 (.37)	.	.
<i>GEBOORT</i> ₋₁	.50 (.094)	.	.	.
Δ <i>OOIVRS</i>	.	.	.	.67 (.21)
Constante	22.07 (3.87)	-15.72 (18.26)	38.37 (3.06)	-1.29 (.95)
Durbin-Watson	2.11	1.38	1.01	1.53
LM-AR(1)	.27 (.62)	1.40 (.28)	.	.
R^2	.96	.50	.91	.53

Tabel 1: Regressieresultaten: Broedparen ooeivaars in Baden-Württemberg (*OOIVRS*) en geboorten in West-Duitsland ($\times 10000$) (*GEBOORT*), 1969-1980 (standaardfouten en overschrijdingskansen)

$Y \rightarrow X$ veroorzaakt. In de definitie verwisselen we dan X en Y . Het is zelfs mogelijk, dat $X \rightarrow Y$ veroorzaakt en $Y \rightarrow X$. Er is dan sprake van simultaneïteit en/of terugkoppeling tussen deze variabelen. We kunnen de gegevens over broedparen ooeivaars in Baden-Württemberg en geboorten in West-Duitsland op Granger causaliteit onderzoeken. We geven deze variabelen aan met *OOIVRS* en *GEBOORT*. De resultaten staan in de tabel.

Uit de eerste twee kolommen blijkt, dat *OOIVRS* een Granger oorzaak is van *GEBOORT*, maar niet andersom. De Lagrange Multiplier toets op eerste-orde autocorrelatie wijst niet op misspecificatie van de vergelijking.¹⁹ Als we de toets op Granger causaliteit met vertraagde waarden van *OOIVRS* uitvoeren, dan is de overschrijdingskans .16 en is *OOIVRS* geen Granger oorzaak van *GEBOORT*. Er is echter geen reden om slechts vertraagde waarden in de relatie op te nemen. Bij de beschouwde relatie gaat het natuurlijk om onmiddellijke causaliteit. De derde kolom bevat de directe regressie van *GEBOORT* op *OOIVRS*. De Durbin-Watson toetsgrootheid is niet veel groter, dan de R^2 en dit is een indicatie van een schijnregressie tussen twee tijdreeksen met een (stochastische) trend. De laatste kolom leert echter, dat ook in eerste verschillen er een significant effect van *OOIVRS* is.

Het gelijkstellen van correlatie aan causaliteit kan dus tot merkwaardige conclusies leiden. We kunnen niet volstaan met het 'laten spreken van de gegevens', te meer omdat analyse van de gegevens kan leiden tot het vinden van schijnverbanden. Deze kunnen echter opgespoord worden door de gevonden 'causale relaties' op verse gegevens te verifiëren. Het probleem van causale inferentie blijft echter bestaan als de gevonden relatie geen schijnrelatie is in deze zin. Denk bijvoorbeeld aan het effect van scholing op inkomen. Dit effect kan zeer stabiel zijn, maar is daarom nog niet causaal.

3 Randomisatie in het Economisch Onderzoek

In tegenstelling tot wat wel eens gedacht wordt, worden er in de economie wel degelijk experimenten uitgevoerd en zelfs experimenten met gerandomiseerde toewi-

¹⁹De gerapporteerde toetsgrootheid is de F -toets van de significantie van het vertraagde residu in de regressievergelijking. In de regressies met vertraagde afhankelijke variabelen is de Durbin-Watson toets onzuiver.

jzing. Ik beperk mij in het vervolg tot veldexperimenten. De bekendste voorbeelden zijn de experimenten met een negatieve inkomstenbelasting, een soort basisinkomen, zoals die in de jaren 60 en 70 in de Verenigde Staten uitgevoerd zijn. In de VS zijn daarnaast economische experimenten uitgevoerd op het gebied van de ziektekostenverzekering, scholing, bonussen voor het vinden van een baan door werklozen, piekbelastingprijzen voor electriciteit en huursubsidie.²⁰

Ik wil mij hier beperken tot de experimenten met scholing, omdat daar de resultaten van experimenten met en zonder gerandomiseerde toewijzing vergeleken zijn. In 1986 publiceerde LaLonde een vergelijking van met verschillende methoden verkregen schattingen van het effect van een scholingsprogramma (LaLonde, 1986). Het scholingsprogramma was de National Supported Work Demonstration (NSW), een experimenteel programma dat tijdelijk banen bood aan bijstandsmoeders, jongeren die hun middelbare school niet afgemaakt hadden, ex-verslaafden en ex-misdadigers. De bedoeling van het programma was om groepen met een zwakke positie op de arbeidsmarkt 9 tot 18 maanden werkervaring in de private of publieke sector te geven, een soort banenpool dus. Na afloop van het programma moesten de deelnemers zelf een andere baan vinden.

De uitvoerders van de NSW besloten het experiment te evalueren door een aantal door het lot aangewezen potentiële deelnemers, d.w.z. personen die aan alle criteria voldeden en wilden meedoen, geen toegang te geven tot de NSW. De toewijzing was derhalve gerandomiseerd. Dit maakte het mogelijk om het gemiddelde causale effect van de NSW op het inkomen 1 jaar na afloop van deelname te meten met het waargenomen verschil tussen het gemiddelde inkomen van de deelnemers en de niet-deelnemers, de controlegroep. Ik concentreer me op het gemeten effect voor de bijstandsmoeders.²¹ Vóór deelname was het gemiddelde inkomen van de bijstandsmoeders in de NSW en de controlegroep nagenoeg gelijk. Dat bevestigt dat de randomisatie goed uitgevoerd is. Tijdens het programma is het inkomen van de deelnemers veel hoger dan dat van de controlegroep. Dat is niet verbazingwekkend, want de deelnemers hebben alle een baan, terwijl een groot deel van de niet-deelnemers nog steeds werkloos is. Het inkomen in de controlegroep stijgt wel. Ook dat is niet vreemd, omdat alle vrouwen op het moment van indeling werkloos waren en ook de vrouwen die niet aan de NSW deelnemen zelf hun positie verbeteren. Eén jaar na afloop van het programma verdienen de ex-NSW deelnemers gemiddeld 22 procent (851 1982 \$) meer dan de niet-deelnemers en dat verschil is in vergelijking met de steekproefvariatie voldoende groot om er enig vertrouwen in te hebben.

LaLonde vergelijkt de schatting die met behulp van gerandomiseerde toewijzing verkregen is met de schatting die verkregen wordt door de NSW deelnemers te vergelijken met een controlegroep die niet door randomisatie verkregen is. Daarvoor gebruikt hij andere gegevensbestanden waarin inkomensgegevens van bijstandsmoeders over de beschouwde periode voorkomen. Vergelijking van het inkomen in het jaar vóór selectie van de gerandomiseerde controlegroep en de uit andere bestanden geconstrueerde controlegroepen leert, dat het niet eenvoudig is een controlegroep te vinden, die bij benadering lijkt op de gerandomiseerde groep. De controlegroep die nog het meest op de gerandomiseerde groep lijkt²², geeft bij vergelijking met de NSW groep een schatting van het causale effect van de NSW, die 3,5 maal zo groot is als het effect dat gevonden is met de gerandomiseerde groep. Opvallend is daarbij dat deze controlegroep op observeerbare kenmerken als leeftijd, school-

²⁰ (Ferber & Hirsch, 1982) geeft een samenvatting.

²¹ De NSW verhoogt het inkomen van mannelijke deelnemers, maar de steekproefvariatie is zo groot voor die groep dat het gemeten effect ook toevallig zou kunnen zijn.

²² Dit is een deelbestand uit de Panel Study of Income Dynamics (PSID) dat bestaat uit bijstandsmoeders tussen 20 en 55 die een bijstandsuitkering ontvingen in 1975 en werkloos waren in 1976. Omdat de NSW alleen op bepaalde plaatsen functioneerde is de kans dat in de controlegroep NSW deelnemers voorkomen verwaarloosbaar.

jaren en ras nauwelijks verschilt van de NSW groep. Het zijn blijkbaar de niet-geobserveerde verschillen tussen de gerandomiseerde en de niet-gerandomiseerde controlegroep die tot de overschatting leiden.

LaLonde onderzoekt ook, of het modelleren van de deelname aan de NSW met het steekproefselectiemodel van Heckman de schattingen die met de niet-gerandomiseerde controlegroep verkregen worden verbetert. De conclusie is, dat het effect nu een factor 2 overschat wordt, en dat een schatting die dichter ligt bij de schatting uit de gerandomiseerde vergelijking alleen verkregen wordt, als de veronderstelling gemaakt wordt, dat variabelen als woonplaats, burgerlijke staat, arbeidsmarktpositie en aantal kinderen alleen de deelname aan de NSW, maar niet het inkomen beïnvloeden.

LaLonde's onderzoek, en ander onderzoek dat LaLonde's resultaten op andere gegevens bevestigde, hebben in de VS tot een hevige discussie geleid.²³ Randomisatie is duidelijk populairder geworden en wordt nu niet alleen bij experimentele, maar ook bij bestaande programma's gebruikt. Het intensievere gebruik van randomisatie heeft echter een aantal problemen aan het licht gebracht. Vooral Heckman heeft in een aantal artikelen de problemen verwoord.²⁴ Deze zijn

- Randomisatie werkt niet bij bestaande programma's. In bestaande programma's gebeurt de selectie door uitvoerende beamten, die daarbij een zekere vrijheid hebben. Zij worden niet graag vervangen door een dobbelsteen en zullen bewust of onbewust de randomisatie saboteren. Een bijkomend probleem is het feit dat er een experiment uitgevoerd wordt invloed kan hebben op de uitkomst.
- Randomisatie werkt alleen in programma's waar er minder plaatsen dan deelnemers zijn. In programma's waar deelnemers zichzelf selecteren is het moeilijk te verkopen, dat iemand niet mee mag doe. De uitgesloten kunnen zich anders gaan gedragen en daarmee een onzuiverheid bij de meting van het causale effect introduceren. Bovendien kan het feit dat er gerandomiseerd wordt potentiële deelnemers afschrikken. Personen die niet mogen meedoen kunnen, zeker als het programma gunstig is, alsnog proberen mee te doen.

Gezien deze problemen is het niet verwonderlijk, dat het tot nu toe moeilijk gebleken is bestaande programma's met gerandomiseerde toewijzing te evalueren. De vraag is, dan ook of er niet een alternatief voor randomisatie bestaat.

4 Indirecte Randomisatie

Randomisatie gaat uit van directe manipulatie van deelname aan een programma. Soms wordt de deelname aan een programma indirect gemanipuleerd. Deze indirecte manipulatie is vaak het onbedoelde gevolg van een interventie, die weliswaar de deelname beïnvloedt, maar verder niets met het programma te maken heeft. We spreken dan van een *natuurlijk experiment*, hoewel er niets natuurlijk aan de interventie behoeft te zijn en de term indirect experiment wellicht meer op zijn plaats is. Om dit begrip te verhelderen behandel ik met u een voorbeeld en wel het onderzoek van Angrist (Angrist, 1990) naar het effect van deelname aan de oorlog in Vietnam op het (levens)inkomen van veteranen.

Directe vergelijking van het inkomen van Vietnam veteranen en niet-veteranen van dezelfde leeftijd geeft een onzuivere schatting van het causale effect. Er was

²³(Heckman & Hotz, 1989) betoogt dat LaLonde met behulp van specificatietoetsen tussen de verschillende schattingen had kunnen discrimineren en in dat geval een schatter gevonden had die niet significant van het 'echte' effect verschilt.

²⁴Onder andere in zijn bijdrage aan een congres dat het Deense Ministerie van Arbeid in 1993 georganiseerd heeft (Heckman & Smith, 1993).

weliswaar in de jaren 1970-1972 dienstplicht en men zou verwachten, dat dit de onzuiverheid zou verkleinen. Echter niet alle mannen die dienstplichtig waren gingen naar Vietnam. Sommigen omdat zij bij keuring ongeschikt bevonden werden, anderen omdat zij over de juiste connecties beschikten om de dienstplicht te ontlopen. Aan de andere kant waren er mannen die dienst namen, hoewel zij niet dienstplichtig waren. Deze onzuiverheden zijn mogelijk groot. Zo wordt voor veteranen uit de tweede wereldoorlog een positief effect op het levensinkomen gevonden en ook sommige studies naar het effect van deelname aan de Vietnamoorlog vinden een positief effect. Hoe kunnen we een zuivere schatting van het gemiddelde causale effect vinden?

Angrist maakt gebruik van een idee van drie medici (Hearst, Newman, & Hulley, 1986). Zij merkten op, dat de selectie voor de dienstplicht op een bijzonder manier gebeurde, namelijk door een loterij. Er zijn tijdens de Vietnamoorlog vier loterijen gehouden, en wel voor de lichtingen 1970 tot en met 1973. De lichting 1973 is uiteindelijk niet opgeroepen. Voor het gemak veronderstellen we, dat de lichtingen geheel uit mannen die 20 jaar eerder geboren zijn bestaan. De loterij voor de lichting 1970 werd op 1 december 1969 gehouden en naar goed Amerikaans gebruik op de televisie uitgezonden. Aan de 365 geboortedata in 1950 werd aselekt een nummer van 1 tot 365 gekoppeld. Vlak voor de datum van opkomst werd een getal vastgesteld, 195 in 1970, en iedereen met een geboortedatum met een lager nummer was dienstplichtig.

Deze procedure randomiseerde de toekenning van de status van dienstplichtige en Hearst, Newman en Hulley gebruikten deze randomisatie om het effect van deelname op de sterftekans van veteranen te bepalen. Zij concludeerden dat het causale effect van deelname aan de Vietnamoorlog op de sterfte in 1979-1984, gemiddeld 5 jaar na afloop van de dienstdienst, 4 procent bedroeg. Ook Angrist maakt gebruik van deze randomisatie van de dienstplichtstatus. Met inkomensgegevens over 1966-1984 voor het cohort dat in 1950 geboren is, constateert hij dat er voor 1970 er geen verschil is tussen het gemiddelde inkomen van dienstplichtigen en vrijgestelden. Om verwarring te voorkomen: ik bedoel met dienstplichtigen mannen met een lotnummer onder de selectiegrens. Dat het verschil gelijk aan 0 is bevestigt de eerlijkheid van de loterij. In 1970 daalt het inkomen van de dienstplichtigen met 234 1978 \$. In de jaren daarna blijft het inkomen van de dienstplichtigen gemiddeld lager.²⁵

Het verschil tussen het gemiddelde inkomen van dienstplichtigen en vrijgestelden kan niet geïnterpreteerd worden als een gemiddeld causaal effect. Immers niet alle dienstplichtigen gingen naar Vietnam en sommige vrijgestelden gingen wel. Het blijkt dat 35 procent van de dienstplichtigen en 19 procent van de vrijgestelden van het cohort 1950 in Vietnam dienden. Om een schatting van het causale effect van deelname aan de oorlog op het latere inkomen te krijgen moeten we het gemeten verschil in gemiddeld inkomen corrigeren voor de beperkte deelname van dienstplichtigen en de deelname van vrijgestelden. Deze correctie kan uitgevoerd worden door het verschil in gemiddeld inkomen te delen door het verschil in deelname tussen beide groepen. Voor het cohort 1950 en het jaar 1981, ongeveer 7 jaar na afloop van de dienstdienst, vinden we zo $-487.8/.159 = -2195.8$ 1978 \$ met een standaardfout van 1069.5. Dit getal is gelijk aan de schatting die verkregen wordt bij een wat efficiënter gebruik van de beschikbare gegevens. Deze schatting correspondeert met een inkomensverlies van 15 procent.

We kunnen de schatting van Angrist interpreteren in het RH model voor causale inferentie. Om dezelfde notatie te kunnen gebruiken moeten we nu deelname aan de Vietnamoorlog als een scholingsprogramma opvatten. Verder introduceren we een indicator variabele Z_i die gelijk is aan d als i dienstplichtig is en gelijk is aan v als i

²⁵ Deze gegevens hebben betrekking op blanke dienstplichtigen. Voor zwarte dienstplichtigen is het verschil eerst ook negatief, maar later positief.

vrijgesteld is. Het gemeten inkomensverschil tussen dienstplichtigen en vrijgestelden is

$$E(Y_S | Z = d) - E(Y_S | Z = v) \quad (6)$$

en dat kan geschreven worden als

$$\begin{aligned} & (E(Y_c | Z = d) - E(Y_c | Z = v)) \\ & + E((Y_s - Y_c)I(S = s) | Z = d) \\ & - E((Y_s - Y_c)I(S = s) | Z = v) \end{aligned} \quad (7)$$

De gerandomiseerde toewijzing van de dienstplichtstatus maakt de eerste term gelijk aan 0. Als we veronderstellen, dat het causale effect $Y_s - Y_c$ voor iedereen gelijk is, dan kunnen we het verschil van de tweede en derde term herschrijven als

$$(Y_s - Y_c)(P(S = s | Z = d) - P(S = s | Z = v)) \quad (8)$$

Hier geeft $P(S = s | Z = d)$ de fractie dienstplichtigen die gediend heeft aan. Als we dit gelijkstellen aan vergelijking 6, vinden we

$$Y_s - Y_c = \frac{E(Y_S | Z = d) - E(Y_S | Z = v)}{P(S = s | Z = d) - P(S = s | Z = v)} \quad (9)$$

Dit is de door Angrist gebruikte schatter. Het is de schatter die door Wald in 1940 voorgesteld is voor het regressiemodel met meetfouten. Het is een voorbeeld van een *instrumentele-variabele* (IV) schatter. We zien dat een IV schatter correspondeert met een experiment met indirecte randomisatie.²⁶ De instrumentele variabele Z in Angrists analyse is de gerandomiseerde indicator van de dienstplichtstatus. Een instrumentele variabele correspondeert met een interventie die wel de deelname beïnvloedt, hier de deelname aan de oorlog in Vietnam, maar niet de uitkomst, hier het levensinkomen. Als er een gerandomiseerd natuurlijk experiment uitgevoerd is, dan kunnen we een instrumentele variabele vinden.

Indirecte randomisatie lijkt een mooi idee. Het is echter alleen bruikbaar als natuurlijke experimenten niet zeldzaam zijn. Om u een indruk te geven, zal ik een aantal voorbeelden van natuurlijke experimenten behandelen. Over de meeste valt veel meer te zeggen dan ik nu kan doen. Sommige zijn omstreken, maar wat in de discussie opvalt, is dat de meningsverschillen vooral betrekking hebben op de vraag of de natuur wel gerandomiseerd heeft. Bij een natuurlijk experiment is de notaris niet aanwezig en dus is het goed, dat collega's het enthousiasme van de onderzoekers in bedwang houden. De natuurlijke experimenten die ik met u zal bespreken, hebben betrekking op de dienstplicht, de immigratie en de invloed van de geboortedatum.

Ook in Nederland is het effect van de dienstplicht op het latere inkomen van dienstplichtigen onderzocht en wel door Imbens en Van der Klaauw (Imbens & Van der Klaauw, 1995). Hoewel Nederland nog steeds algemene dienstplicht kent, dient minder dan de helft van de in een bepaald jaar geboren mannen. De fractie van een geboortecohort die dient, varieert sterk. De in 1959 geboren mannen werden zelfs collectief vrijgesteld van militaire dienst. De variatie in de fractie wordt vooral bepaald door de vraag naar diensplichtigen en de veronderstelling dat het Ministerie van Defensie daarbij vooral aan zichzelf denkt, is redelijk. Dit betekent dat de fractie van een geboortecohort die uiteindelijk dient, een instrumentele variabele is, die wel invloed heeft op het al of niet dienen, maar niet op het latere inkomen. Gebruik makend van dit idee vinden Imbens en Van der Klaauw dat dienstplicht het latere jaarinkomen met 8 procent doet dalen. Dit correspondeert met het verlies van twee jaar werkervaring, meer dan de 14 maanden die de dienst gemiddeld duurt.

²⁶In de afleiding heb ik verondersteld, dat het causale effect voor iedereen gelijk is. (Imbens & Angrist, 1994) bespreekt wat de IV schatter schat als dat niet het geval is.

Het tweede voorbeeld betreft het effect van immigratie op de arbeidsmarktpositie van autochtonen. Op 20 april 1980 verklaarde Fidel Castro, dat iedereen die Cuba wilde verlaten, uit de havenstad Mariel mocht vertrekken. Dit leidde in mei tot september 1980 tot een uittocht van Cubanen, waarvan ongeveer 125000 in Miami en omgeving terecht kwamen. Hierdoor groeide het arbeidsaanbod in Miami met 7 procent. Dat van Cubanen zelfs met 20 procent. Hoewel deze interventie voor de VS als een verrassing kwam en dus opgevat kan worden als een natuurlijk experiment, wordt de analyse gecompliceerd doordat de effecten over de tijd zichtbaar werden. Een simpele voor-na vergelijking kan vertekend worden door veranderingen in het niveau van de bedrijvigheid. Vergelijking met andere steden lijkt daarom gewenst. Card (Card, 1990) heeft de Mariel immigratie geanalyseerd. Een belangrijke conclusie is dat de lonen van zwarte werknemers, die vaak als concurrenten van de Cubanen op de arbeidsmarkt gezien worden, na 1981 begonnen te dalen, maar dat zij in 1985 weer op het oude niveau terug waren. In andere steden daalden de lonen van zwarte werknemers na 1981 ook, maar trad geen herstel op. Het werkloosheidspercentage onder zwarte werknemers steeg in 1981 en volgende jaren sterker dan in andere steden, maar daalde na 1983 weer. Hetzelfde patroon deed zich voor bij Cubanen, terwijl de werkloosheid van blanken in Miami getrouw de ontwikkeling in andere steden volgde. Bij de beoordeling van deze resultaten blijkt een beperking van natuurlijke experimenten: de generalisatie. Resultaten voor Miami met een traditie van immigratie en een daarmee corresponderende bedrijvigheid hoeven niet op te gaan voor andere steden in andere landen, maar het resultaat geeft toch te denken.²⁷

Zijn er in Nederland gegevens die als de uitkomst van een natuurlijk experiment gezien kunnen worden? Wie de ontwikkelingen op het gebied van het vreemdelingenrecht en het beleid met betrekking tot de hier ruim getrokken kring van allochtonen volgt, vermoedt dat willekeurige veranderingen en beperkingen vele natuurlijke experimenten genereren. Ik wil hier op de verrassende uitkomst van één studie wijzen, maar tegelijk aangeven dat nadere analyse nodig is om er zeker van te zijn, dat het hier om een natuurlijk experiment gaat. In een tweetal rapporten heeft het onderzoeksbureau Regioplan de positie van twee groepen vreemdelingen onderzocht.²⁸ De groepen zijn vreemdelingen die in 1991 erkend zijn als vluchteling met een A-status of een vergunning tot verblijf hebben ontvangen en vreemdelingen die in 1992 gedoogd werden en op grond van artikel 84 van de Algemene Bijstandswet een beroep op de bijstand konden doen. De eerste groep vreemdelingen kreeg voor erkenning een uitkering en huisvesting op grond van de Regeling Opvang Asielzoekers (ROA), terwijl de tweede groep daarvan uitgesloten was. Vergelijking van de twee groepen wordt bemoeilijkt, doordat de tweede groep al enige tijd in Nederland verbleef, gemiddeld ongeveer 3 jaar. Om die reden moeten we de erkenden lange tijd volgen om een vergelijking met de gedooften te kunnen maken. Bij vergelijking van de fractie personen met een bijstandsuitkering blijkt, dat 12 procent van de gedooften (1955 waarnemingen) en 80 procent (543 waarnemingen) van de erkenden een bijstandsuitkering ontvangt. Het laatste percentage heeft betrekking op de situatie na 2 jaar en is daarom strikt genomen niet vergelijkbaar met de fractie voor de gedooften. Voordat we echter de conclusie trekken, dat de ROA de positie van vreemdelingen nadelig beïnvloedt, moeten we meer weten over de 'toevalligheid' van de toelating tot de ROA. Helaas richt het onderzoek zich uitsluitend op de kenmerken van de ontvangers van een uitkering en is er niets bekend over de niet-ontvangers, zodat zelfs een vergelijking van de kenmerken van erkenden en gedooften onmogelijk is.

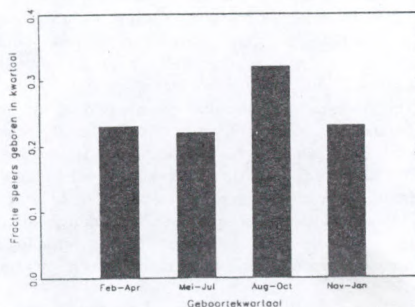
Het derde voorbeeld heeft betrekking op het effect van de geboortemaand op

²⁷ Zie ook de analyse in (Hunt, 1992) van het effect van de repatriëring van Fransen uit Algerije.

²⁸ (Visser, Van Waveren, & Homburg, 1994; Van Waveren, Visser, & Homburg, 1994)

het succes in het leven, bekend terrein voor astrologen. In de meeste landen gaan kinderen niet naar school op de dag nadat zij zes jaar geworden zijn. In de Verenigde Staten geldt dat een kind alleen tot de eerste klas toegelaten wordt, als het vóór 1 januari van het schooljaar zes wordt. De leerplicht houdt echter precies op de zestiende verjaardag op. Het blijkt nu dat als gevolg hiervan het gemiddeld aantal schooljaren toeneemt met het geboortekwartaal. Als we ervan uitgaan, dat er gemiddeld geen verschil bestaat tussen personen die in verschillende kwartalen geboren zijn, dan is het geboortekwartaal een instrumentele variabele, die wel het aantal schooljaren maar niet direct het latere inkomen beïnvloedt. Angrist en Krueger (Angrist & Krueger, 1991) hebben dit natuurlijke experiment gebruikt om de opbrengst van onderwijs te schatten. Zij vonden dat traditionele methoden er niet ver naast zaten.²⁹

In een ingezonden brief naar het tijdschrift *Nature* laat Dudink (Dudink, 1994) zien, dat er een verband bestaat tussen de geboortemaand en succes in de sport. Hij constateert dat voetballers die in augustus-oktober geboren zijn, oververtegenwoordigd zijn in de Nederlandse professionele competities en dat spelers die in mei-juli geboren zijn sterk ondervertegenwoordigd zijn.³⁰ Hij verklaart deze verdeling uit de indeling van de teams aan het begin van de competitie. Relatief oudere spelers doen het beter en verdringen de relatief jongere. We kunnen de hypothese van Dudink toetsen met een natuurlijk experiment. Op het zuidelijk halfrond begint de competitie niet in september maar in februari. In Brazilië worden de selecties daarom eind januari samengesteld. In december 1994 heb ik in Brazilië gegevens verzameld over de geboortemaand van beroepsvoetballers. Het onderzoek is nog niet afgesloten, maar ik kan u wat voorlopige resultaten geven. Mijn gegevens hebben betrekking op 130 spelers van 13 clubs.³¹ Het gaat hier om droomelftallen, die door liefhebbers samengesteld zijn uit de spelers die in het verleden voor de clubs uitgekomen zijn. Omdat het hier om de beste spelers gaat, verwachten we dat spelers die in februari-april geboren zijn oververtegenwoordigd zullen zijn.



Figuur 2: Verdeling Geboortedata van Voetballers over Kwartalen: Droomelftallen in Brazilië

²⁹Dit heeft geleid tot een herbezinning op IV schatters met een groot aantal en bovendien zwakke instrumenten. (Bound, Jaeger, & Baker, 1993) betoogt, dat IV schatters in dat geval onzuiver zijn. Zie ook (Bekker, 1994).

³⁰Dudink heeft de geboortemaand van 621 spelers van 36 clubs.

³¹Placar, november 1994.

Uit de figuur blijkt dat dit niet het geval is.^{32,33} Het lijkt erop dat Dudink's hypothese niet bestand is tegen dit natuurlijke experiment.

Tenslotte wil ik u twee voorbeelden geven van (nog) niet uitgevoerde indirecte experimenten. Het eerste heeft betrekking op de bijstand. Sinds enige tijd probeert de overheid te bevorderen dat ontvangers van een bijstandsuitkering weer aan het werk gaan.³⁴ De contactpersonen op de Gemeentelijke Sociale Diensten, de bijstandsmagazijnhouders (BMW), beschikken daartoe over een aantal positieve en negatieve prikkels. De BMW's hebben ook veel vrijheid bij het toepassen van deze prikkels, zodat een directe meting mogelijk een onzuivere schatting van het effect geeft. Directe randomisatie lijkt ook niet goed mogelijk. De koppeling van cliënten en BMW's is echter min of meer toevallig en kenmerken van de BMW, zoals gegevens over de prikkels die hij of zij in het afgelopen jaar toegediend heeft, kunnen als instrumentele variabele gebruikt worden. Het tweede indirecte experiment heeft betrekking op de Individuele Huursubsidie. Ruud Koning en ik hebben op niet-experimentele wijze het effect van de IHS op de woonconsumptie geschat. Er is ook hier een indirect experiment mogelijk. Het blijkt namelijk, dat niet ieder rechtgebend huishouden IHS aanvraagt. Door nu aselect een aantal huishoudens attent te maken op hun recht creëren we een indirect experiment.

Natuurlijke of indirecte experimenten zijn niet alomtegenwoordig, maar met enige inspanning vinden we ze op onverwachte plaatsen.

5 Experimenten en Economische Theorie

Tenslotte wil ik nog een paar opmerkingen maken over de rol van de economische theorie in het onderzoek naar causale effecten. Uit het voorgaande zou de indruk kunnen ontstaan dat die niet groot is. Ik zou daarmee mijn eigen onderzoek, waarin ik die theorie veel gebruik, verloochenen. Als ik het over de economische theorie heb, dan doel ik niet op een verzameling formele theorieën, maar meer op de basisgedachte van de economische wetenschap, dat beslissers gegeven de beperkingen waarmee zij geconfronteerd worden, er het beste van proberen te maken. Op basis van dit inzicht kon de ondergang van het systeem van centrale planning met zijn merkwaardige incentives al bij invoering voorspeld worden. Het is de tragiek van de theorie, dat het vervolgens 70 jaar duurde voor het zover was.

De economische theorie kan de zoeker naar causale effecten op twee manieren helpen en wel door *identificatie* en door *interpretatie*. Een goed voorbeeld van het eerste is het Roy model. Heckman en Honoré (Heckman & Honoré, 1990) laten zien, dat als de beslissers de beste optie kiezen, we een zuivere schatting van het effect kunnen krijgen, zonder allerlei verdelingsveronderstellingen op te hoeven leggen. We bevinden ons dan op het gebied van de niet-parametrische identificatie. Opmerkelijk genoeg veronderstellen we bij dit soort resultaten, dat de verdeling van de niet-observeerbare determinanten onafhankelijk is van de observeerbare grootheden. Dus toch weer randomisatie?

Economische theorie is ook nodig bij de interpretatie van de gevonden effecten. Zonder interpretatie vernauwt onze kennis zich tot een aantal weliswaar zuiver gemeten, maar verder ongerelateerde causale relaties. Theorieën die consistent zijn met die relaties suggeren nieuwe metingen en daarmee nieuwe inzichten. En zoals de uitvinder van de experimentele statistiek Fisher al opmerkte: als er geen experimenten mogelijk zijn, wees dan expliciet over je theorie³⁵ en ga daarbij in ieder geval na of je theorie consistent is met bekende feiten. Als directe of indirecte experi-

³²De hypothese dat de verdeling uniform is wordt niet verworpen. De overschrijdingskans is .29.

³³Voor de liefhebbers: Romário is op 29 januari geboren, Pelé op 23 november

³⁴(Angenent, Bommelijé, & Schep, 1993)

³⁵(Cochran, 1965)

iminten niet mogelijk zijn, dan zijn we aangewezen op wat Diaconis (Diaconis, 1985) beheerste magie noemt. De gevolgen van interventies worden voorspeld op basis van theorieën, die voor een belangrijk deel speculatief zijn, natuurlijk zo min mogelijk. De econoom en de econometrist lijken dan ook wel wat op medicijnmannen. Die staan overigens in hoog aanzien bij stam.

Hoe zat het nu met die Filistijnen? Die dachten de oorzaak van Simsons kracht ontdekt te hebben en organiseerden een groot feest waar ook de getemde krachtpatser moest opdraven. Het vertrouwen in hun causale inferentie was echter te groot en in een zelfmoordactie deed Simson de tempel waarin het feest gehouden werd instorten. Als u goed opgelet heeft, weet u waar de Filistijnen in de fout gingen. Haastige conclusies zijn soms dodelijk.

6 Literatuur

- Angenent, F. J. A., Y. B. Bommeljé, and G. J. Schep. *Prikkels in de bijstand*. 's-Gravenhage: Ministerie van Sociale Zaken en Werkgelegenheid/VUGA, 1993.
- Angrist, Joshua D. "Lifetime earnings and the Vietnam era draft lottery: evidence from social security administrative records." *American Economic Review* 80, no. 3 (1990): 313-36.
- Angrist, Joshua D., and Alan B. Krueger. "Does compulsory school attendance affect schooling and earnings?" *Quarterly Journal of Economics* CVI, no. 4 (1991): 979-1014.
- Ashenfelter, Orley. "Estimating the effect of training programs on earnings." *Review of Economics and Statistics* 60 (1978): 47-57.
- Bekker, Paul A. "Alternative approximations to the distributions of instrumental variable estimators." *Econometrica* 62, no. 3 (1994): 657-82.
- Van den Berg, G.J. "Nonstationarity in job search theory" *Review of Economic Studies* 57 (1990): 255-77.
- Black, Fisher. "The trouble with econometric models." *Financial Analysts Journal* (March-April 1982): 29-37.
- Bound, J., D. Jaeger, and R. Baker. The cure can be worse than the disease: A cautionary tale regarding instrumental variables. mimeo, University of Michigan.
- Bouter, L. M. Extramurale epidemiologie: gewoon en bijzonder. Oratie, Faculteit der Geneeskunde, Vrije Universiteit.
- Card, David. "The impact of the Mariel boatlift on the Miami Labor Market." *Industrial and Labor Relations Review* 43, no. 2 (1990): 245-57.
- Cochran, W. G. "The planning of observational studies of human populations." *Journal of the Royal Statistical Society, A* 182 (1965): 234-55.
- Diaconis, Persi. "Theories of data analysis: From magical thinking through classical statistics." In *Exploring Data Tables, Trends, and Shapes*, edited by David C. Hoaglin, Frederick Mosteller, and John W. Tukey, 1-36. New York: John Wiley & Sons, 1985.
- Dudink, Ad. "Birth date and sporting success." *Nature* 368 (14 April 1994): 592.
- Ferber, Robert, and Werner Z. Hirsch. *Social Experimentation and Economic Policy*. New York: Cambridge University Press, 1982.
- Fisher, R. A. "The arrangement of field experiments." *Journal of Ministry of Agriculture* 33 (1926): 503-13.

- Fréchet, M. "Sur les tableaux de corrélation dont les marges sont données." *Annales de Université, Lyons Sect A* 14 (1951): 53-77.
- Freedman, David. From association to causation via regression. Technical report no. 408, Statistics Department, University of California, Berkeley.
- Glymour, C. et al. *Discovering Causal Structure*. New York: Academic Press, 1993.
- Granger, C. W. J. "Investigating causal relations by econometric models and cross-spectral methods." *Econometrica* 37, no. 3 (1969): 424-38.
- Haavelmo, T. "The probability approach to econometrics." *Econometrica* 12, no. Supplement (1944).
- Hearst, Norman, Thomas B. Newman, and Stephen B. Hulley. "Delayed effect of the military draft on mortality: A randomized natural experiment." *New England Journal of Medicine* 314 (6 March 1986): 620-24.
- Heckman, James J., and Bo Honoré. "The empirical content of the Roy model." *Econometrica* 58 (1990): 1121-49.
- Heckman, James J., and V. Joseph Hotz. "Choosing among alternative nonexperimental methods for estimating the impact of social programs: The case of manpower training." *Journal of the American Statistical Association* 84 (1989): 862-80.
- Heckman, James J., and Jeffrey Smith. "Assessing the case for randomized evaluation of social programs." In *Measuring labour market measures*, edited by Karsten Jensen and Per Kongshoj Madsen, 35-96. Copenhagen: Ministry of Labour, 1993.
- Heckman, James, J. "Sample selection bias as a specification error." *Econometrica* 47 (1979): 153-61.
- Hendry, David F. *Dynamic Econometrics*. Oxford: Oxford University Press, 1995.
- Holland, Paul W. "Statistics and causal inference." *Journal of the American Statistical Association* 81 (1986): 945-70.
- Hunt, Jennifer. "The impact of the 1962 repatriates from Algeria on the French labor market." *Industrial and Labor Relations Review* 45, no. 3 (1992): 556-72.
- Imbens, G., and J. Angrist. "Identification and estimation of local average treatment effects." *Econometrica* 62, no. 2 (1994): 467-76.
- Imbens, Guido, and Wilbert Van der Klaauw. "Evaluating the cost of conscription in the Netherlands." *Journal of Business and Economic Statistics* forthcoming (1995).
- Keynes, J. M. "Professor Tinbergen's method." *Economic Journal* 49 (1939): 558-68.
- LaLonde, Robert J. "Evaluating the econometric evaluations of training programs with experimental data." *American Economic Review* 76, no. 4 (1986): 604-20.
- Morgan, Mary S. *The History of Econometric Ideas*. Cambridge: Cambridge University Press, 1990.
- Roy, A. D. "Some thoughts on the distribution of earnings." *Oxford Economic Papers* 3 (1951): 135-46.
- Rubin, D. B. "Estimating causal effects of treatments in randomized and nonrandomized studies." *Journal of Educational Psychology* 66 (1974): 688-701.
- Sies, Helmut. "A new parameter in sex education." *Nature* (7 April 1988): 495.
- Spirtes, P. et al. TETRAD II. Technical report, Department of Philosophy, Carnegie Mellon University, Pittsburgh.
- Suppes, P. *A Probabilistic Theory of Causality*. Amsterdam: North-Holland, 1970.

- Visser, E. T., R. C. Van Waveren, and G. H. J. Homburg. *Vreemdelingen en bijstand, deel I: Artikel-84-vreemdelingen in de bijstand*. Raport Regioplan B.V.
- Wang, C. *Sense and Nonsense of Statistical Inference*. New York: Marcel Dekker, Inc., 1993.
- Van Waveren, R. C., E. T. Visser, and G. H. J. Homburg. *Vreemdelingen en bijstand, deel II: Ex-asielzoekers in de bijstand*. 's- Gravenhage: Ministerie van Sociale Zaken en Werkgelegenheid/VUGA, 1994.

